KMITL Sci. J. Vol. 5 No. 1 Feb. 2005

On a Comparison of two Standard Estimators of a Binomial Proportion by Multiple Criteria Decision Making Method

Montip Tiensuwan[*,1], Satinee Lertprapai[*,2], and Bimal K Sinha[**]

[*]Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand

[**]Department of Mathematics and Statistics, University of Maryland, Baltimore County, USA

Abstract

In this paper, we consider the problem of estimation of a binomial proportion θ based on $X \sim B(n,\theta)$, n being known, $0<\theta<1$, θ being unknown. We compare two standard estimators of θ, namely, $T_1 = X/n$, and $T_2 = [X+\sqrt{n}/2]/[n+\sqrt{n}]$, on the basis of Multiple Criteria Decision Making (MCDM) procedure. Our recommendation is that we should use T_2 rather than T_1 for small values of n.

Keywords: Binomial distribution, proportion, multiple criteria decision making

---

[1] Research supported by a grant from The Thailand Research Fund (TRF).

[2] Research supported by a scholarship under the Commission on Higher Education, Staff Development Project, Thailand.

E-mail: scmts@mahidol.ac.th

1. Introduction

We consider the problem of estimation of a binomial proportion θ based on $X \sim B(n,\theta)$, n being known, $0<\theta<1$, θ being unknown. We compare two standard estimators of θ, namely, $T_1 = X/n$, the maximum likelihood estimate [1] and $T_2 = [X+\sqrt{n}/2]/[n+\sqrt{n}]$, the minimax estimate under the squared error loss function [1]. In this paper we compare T_1 and T_2 on the basis of Multiple Criteria Decision Making (MCDM) method. This method is briefly described in Section 2 and Section 3 contains the main results of this paper. Our recommendation is to use T_2 rather than T_1 for most reasonable values of n on binomial proportion. For detailed discussion on MCDM, we refer to Zeleny[2].

2. Multiple Criteria Decision Making procedure (MCDM)

2.1 Introduction of MCDM

Multiple Criteria Decision Making (MCDM) has recently been recognized as an efficient statistical method to combine component 'indices' arising from many 'sources' into a single overall meaningful index. Such an index can be effectively used to compare relevant 'facilities'. The basic premise is a data matrix $X = (x_{ij}) : K \times N$ where the rows represent facilities which need to be compared or ranked with respect to the element x_{ij}'s, the columns represent various sources of the elements x_{ij}'s and x_{ij}'s themselves represent some quantitative information about the facilities. In the context of environmental science, the x_{ij}'s may represent levels of pollutants, facilities represent the sources of the pollutants (e.g., chemical or nuclear facilities) and the columns represent different types of pollution. Since usually it is difficult to compare the facilities on a multiple scale, MCDM provides a statistical method to combine the elements in any row into a single value which can then be used to compare the rows on a linear scale.

MCDM is a procedure to integrate multiple indicators into a single meaningful and overall index by combining $(x_{i1},\ldots,x_{iN})$ for row i across all indicators $j=1,2,\ldots,N$. We can define an Ideal Row as one with the smallest observed value for each column

$$IDR = (min_i\, x_{i1},\ldots,min_i\, x_{iN}) = (u_1,\ldots,u_N)$$

and a Negative-ideal Row (NIDR) as one with the largest observed value for each column

$$NIDR = (max_i x_{i1}, \ldots, max_i x_{iN}) = (v_1, \ldots, v_N)$$

For any given row i, we now compute the distance of each row from Ideal row and from Negative Ideal row based on a suitably chosen norm. Under L_1-norm, we compute

$$L_1(i, IDR) = \sum_{j=1}^{N} \frac{|x_{ij} - u_j| w_j}{\sum_{i=1}^{K} x_{ij}} = \sum_{j=1}^{N} \frac{[x_{ij} - u_j] w_j}{\sum_{i=1}^{K} x_{ij}},$$

$$L_1(i, NIDR) = \sum_{j=1}^{N} \frac{|x_{ij} - v_j| w_j}{\sum_{i=1}^{K} x_{ij}} = \sum_{j=1}^{N} \frac{[v_j - x_{ij}] w_j}{\sum_{i=1}^{K} x_{ij}}$$

where $w_1, w_2, \ldots, w_N$ are suitably chosen nonnegative weights between 0 and 1. The denominator above plays the role of a '*norming*' factor. An objective way to select the weights is to use Shannon's [5] entropy measure ϕ based on the proportion $p_{1j}, \ldots, p_{Kj}$ for the jth column where

$$p_{ij} = x_{ij} / \sum_{i=1}^{K} x_{ij}$$

For the jth column, ϕ_j is computed as

$$\phi_j = -\sum_{i=1}^{K} p_{ij} \ln(p_{ij}) \; / \; [\ln(K)].$$

The quantity ϕ essentially provides a measure of closeness of the different proportions. The smaller the value of ϕ, the larger the variation among the proportions for classifying the rows. So we can select the weights as

$$w_j = (1-\phi_j) \; / \; [\sum_{j=1}^{N} (1-\phi_j)] \; , \; j = 1, \ldots, N.$$

In addition to Shannon's entropy measure, we can also use the sample variance of these proportions, given by

$$s_{j\;prop}^2 = \sum_{j=1}^{N} (p_{ij} - \bar{p}_j)^2 / (K-1).$$

If $\bar{x}_j$ and s_j^2 denote the mean and variance of x_{ij} in the j th column , $s^2_{j/prop}$ is directly proportional to $s_j^2 / \bar{x}_j^2$, which is the square of the sample coefficient of variation cv_j. Therefore we propose to use $w_j = cv_j$.

The various rows are now compared based on an overall index computed as

$$L_1(Index_i) = \frac{L_1(i, IDR)}{L_1(i, IDR) + L_1(i, NIDR)}, \quad i = 1, \ldots, K. \qquad (2.1)$$

Similarly, under L_2-norm, we compute

$$L_2(i, IDR) = [\sum_{j=1}^{N} (x_{ij} - u_j)^2 w_j]^{1/2}$$

$$L_2(i, NIDR) = [\sum_{j=1}^{N} (x_{ij} - v_j)^2 w_j]^{1/2}$$

The various rows are now ranked based on an overall index I computed as

$$L_2(Index_i) = \frac{L_2(i, IDR)}{L_2(i, IDR) + L_2(i, NIDR)}, \quad i = 1, \ldots, K. \qquad (2.2)$$

2.2 An application

In this section we apply the previously described MCDM method to the air pollution data from Bangkok, Thailand for our numerical example. The main air pollutants in Bangkok are carbon monoxide (CO_2), nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) which are released directly from motor vehicles. The photochemical reaction on the oxide of nitrogen is ozone (O_3) which is a secondary pollutant.

The data sets were provided by the Pollution Control Department of Thailand and were recorded by 10 monitoring stations in Bangkok during 1998 – 2001. The monitoring stations are as follows:

1. Ramkhamheang University
2. National Housing Authority
3. Huai Khwang
4. Nonsee Vitaya School
5. Singharatpitayakom School
6. Thonburi
7. Chokchai 4
8. Dindaeng
9. Meteorological Department
10. Ratburana.

The locations of 10 monitoring stations in Bangkok are shown in Figure 1.

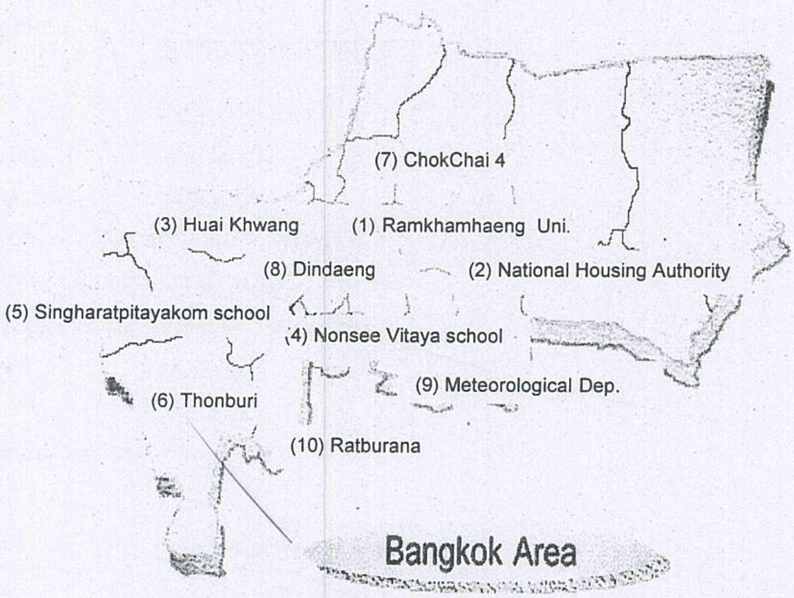

Figure 1 : Location of 10 monitoring stations in Bangkok area

At each station, the signals from the instruments were sampled every five seconds and hourly average values were calculated and stored. For our analysis, we have used the annual averages of each pollutant. The entire data set appears in a Technical Report (Lertprapai et al.[3]).

To apply the MCDM method, we use both the distance measures L_1 and L_2 as well as the two choices of weights based on phi and coefficient of variation (cv). We show below the results in four sets of the values of combined indices for each year. The final ranks of the rows are then based on the average index. We also compute the standard deviation to show the closeness of the four indices in a row. Tables 1 – 4 present all the results for years 1998 - 2001.

From Tables 1-4, we observe that most often (1998, 2000, 2001) first rank is Meteorological Department station which means this station is expected to be good in terms of air pollution. On the other hand, Dindaeng station performed poorly. We selected these two stations to represent their performances graphically in Figures 2 – 3. These figures also depict their ranks for each season separately, rainy, summer and winter, along with the overall ranks. Details of seasonal analyses appear in the Technical Report [3].

Table 1 : Results of MCDM method on air pollution data in 1998.

Monitoring station	L_1		L_2		Mean	SD	rank
	W_1	W_2	W_1	W_2			
(1) Ramkhamheang University	0.3574	0.3610	0.3891	0.3922	0.3749	0.0183	6
(2) National Housing Authority	0.2983	0.3023	0.3327	0.3359	0.3173	0.0197	4
(3) Huai Khwang	0.4423	0.4461	0.4390	0.4425	0.4425	0.0029	9
(4) Nonsee Vitaya school	0.2934	0.2896	0.3271	0.3235	0.3084	0.0196	3
(5) Singharatpitayakom school	0.3707	0.3685	0.3932	0.3937	0.3815	0.0138	7
(6) Thonburi	0.4255	0.4293	0.4281	0.4316	0.4286	0.0025	8
(7) Chokchai 4	0.3685	0.3665	0.3716	0.3698	0.3691	0.0021	5
(8) Dindaeng	0.8054	0.7983	0.6909	0.6855	0.7450	0.0657	10
(9) Meteorological Department	0.0387	0.0402	0.0492	0.0503	0.0446	0.0060	1
atburana	0.1217	0.1236	0.1828	0.1858	0.1535	0.0356	2

Table 2 : Results of MCDM method on air pollution data in 1999.

Monitoring station	L_1		L_2		Mean	SD	rank
	W_1	W_2	W_1	W_2			
(1) Ramkhamheang University	0.4414	0.4374	0.4683	0.4667	0.4534	0.0163	9
(2) National Housing Authority	0.3281	0.3154	0.3709	0.3634	0.3445	0.0269	3
(3) Huai Khwang	0.4271	0.4172	0.4287	0.4190	0.4230	0.0058	7
(4) Nonsee Vitaya school	0.3519	0.3367	0.3745	0.3621	0.3563	0.0160	4
(5) Singharatpitayakom school	0.4193	0.4004	0.4328	0.4169	0.4174	0.0133	6
(6) Thonburi	0.4499	0.4451	0.4464	0.4411	0.4456	0.0036	8
(7) Chokchai 4	0.3654	0.3609	0.3665	0.3615	0.3636	0.0028	5
(8) Dindaeng	0.7386	0.7341	0.6180	0.6146	0.6763	0.0694	10
(9) Meteorological Department	0.1890	0.1984	0.2448	0.2518	0.2210	0.0319	2
(10) Ratburana	0.1382	0.1404	0.1831	0.1851	0.1617	0.0259	1

Table 3 : Results of MCDM method on air pollution data in 2000.

Monitoring station	L_1		L_2		Mean	SD	rank
	W_1	W_2	W_1	W_2			
(1) Ramkhamheang University	0.4031	0.4041	0.4141	0.4150	0.4091	0.0064	9
(2) National Housing Authority	0.2499	0.2470	0.2951	0.2930	0.2713	0.0264	4
(3) Huai Khwang	0.3090	0.3190	0.3038	0.3126	0.3111	0.0064	7
(4) Nonsee Vitaya school	0.2067	0.2305	0.2485	0.2660	0.2379	0.0254	2
(5) Singharatpitayakom school	0.3392	0.3315	0.3600	0.3558	0.3466	0.0135	8
(6) Thonburi	0.2907	0.2958	0.2909	0.2948	0.2930	0.0026	5
(7) Chokchai 4	0.3010	0.3143	0.2929	0.3045	0.3032	0.0089	6
(8) Dindaeng	0.7350	0.7544	0.6410	0.6489	0.6948	0.0582	10
(9) Meteorological Department	0.1420	0.1320	0.1599	0.1548	0.1472	0.0126	1
(10) Ratburana	0.2370	0.2243	0.3049	0.2983	0.2661	0.0414	3

Table 4 : Results of MCDM method on air pollution data in 2001.

Monitoring station	L_1		L_2		Mean	SD	rank
	W_1	W_2	W_1	W_2			
(1) Ramkhamheang University	0.4144	0.3669	0.4595	0.4356	0.4191	0.0394	8
(2) National Housing Authority	0.3074	0.3006	0.3512	0.3411	0.3250	0.0248	7
(3) Huai Khwang	0.2739	0.2891	0.2730	0.2850	0.2803	0.0081	6
(4) Nonsee Vitaya school	0.1967	0.2082	0.2155	0.2226	0.2107	0.0111	3
(5) Singharatpitayakom school	0.4501	0.4371	0.4554	0.4414	0.4460	0.0083	9
(6) Thonburi	0.1853	0.2000	0.1865	0.1991	0.1927	0.0079	2
(7) Chokchai 4	0.2237	0.2342	0.2238	0.2331	0.2287	0.0057	4
(8) Dindaeng	0.5948	0.6508	0.5424	0.5680	0.5890	0.0464	10
(9) Meteorological Department	0.1684	0.1544	0.1954	0.1852	0.1759	0.0181	1
(10) Ratburana	0.2131	0.1946	0.2675	0.2539	0.2323	0.0341	5

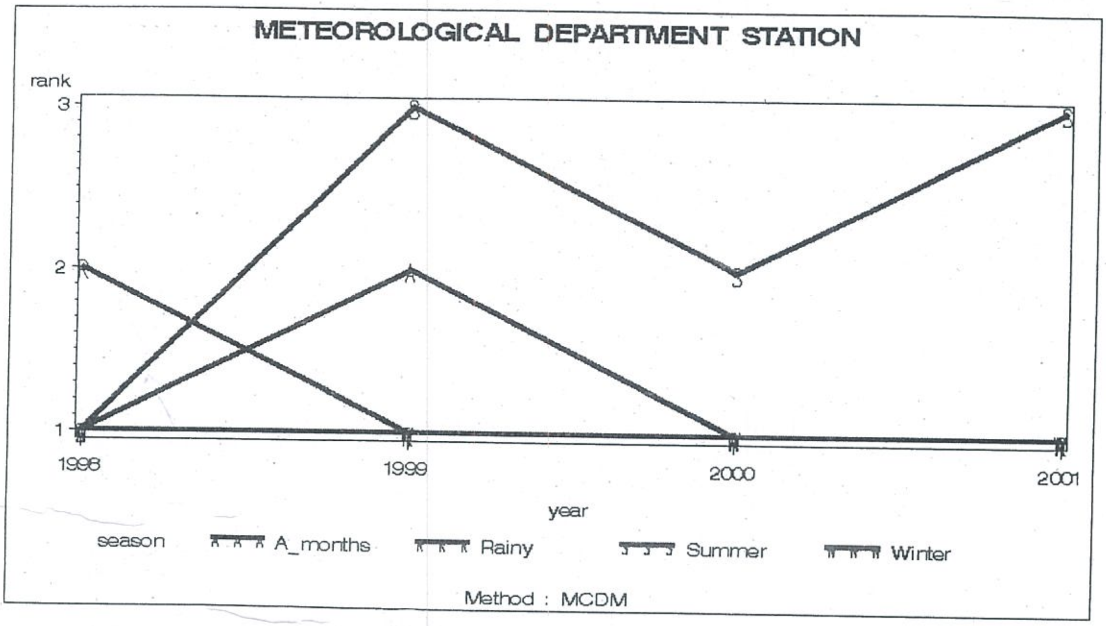

Figure 2 : Order of rank of Meteorological Department station for 1998-2001.

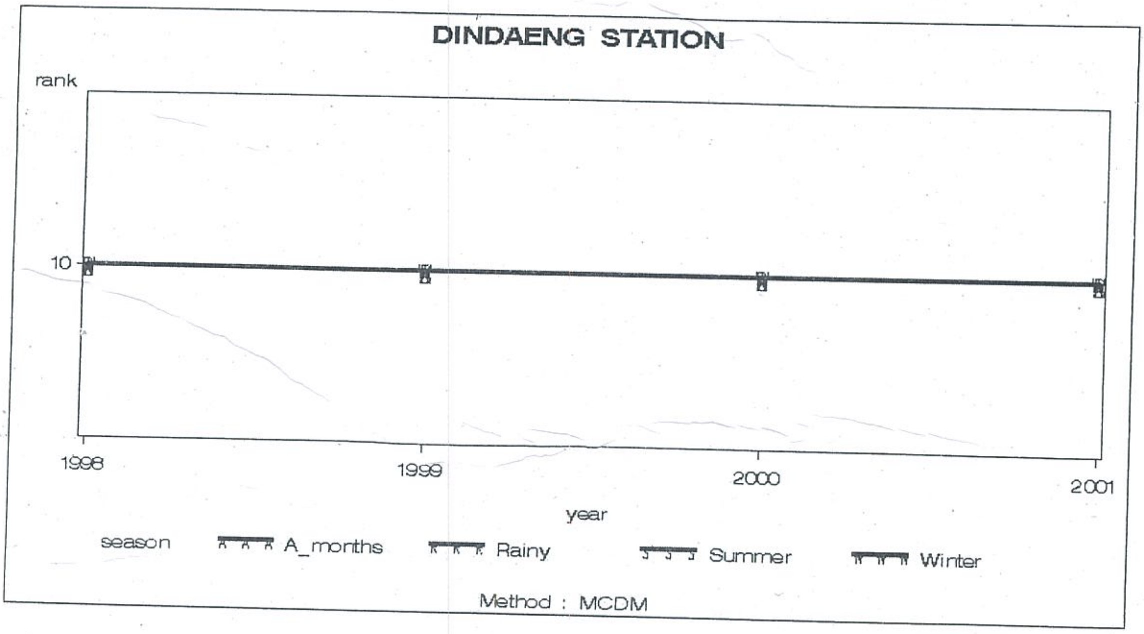

Figure 3 : Order of rank of Dindaeng station for 1998-2001.

2.3 Binomial proportion

According to a '*continuous*' version of this setup would involve $x_{ij}'s$ where the index j would vary '*continuously*'. In the context of our problem of comparing T_1 and T_2 for estimation of θ, obviously $K = 2$, $x_{ij}'s$ are chosen to represent the mean squared errors of T_1 and T_2 for various values of θ, and L_1-norm and L_2-norm would be redefined as

$$L_l(i, IDR) = \int_{\underline{\theta}}^{\overline{\theta}} [x_i(\theta) - u(\theta)] w(\theta)\, d\theta \tag{2.3}$$

$$L_l(i, NIDR) = \int_{\underline{\theta}}^{\overline{\theta}} [v(\theta) - x_i(\theta)] w(\theta)\, d\theta \tag{2.4}$$

$$L_2(i, IDR) = \sqrt{\int_{\underline{\theta}}^{\overline{\theta}} (x_i(\theta) - u(\theta))^2 w(\theta)\, d\theta} \tag{2.5}$$

$$L_2(i, NIDR) = \sqrt{\int_{\underline{\theta}}^{\overline{\theta}} (x_i(\theta) - v(\theta))^2 w(\theta)\, d\theta} \tag{2.6}$$

where $u(\theta) = \min_i \{x_i(\theta)\}$, $v(\theta) = \max_i \{x_i(\theta)\}$, and $\underline{\theta} \le \theta \le \overline{\theta}$.

Comparison of estimates is then based on the overall index defined as

$$L_1(Index_i) = \frac{\int_{\underline{\theta}}^{\overline{\theta}} [x_i(\theta) - u(\theta)] w(\theta)\, d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} [x_i(\theta) - u(\theta)] w(\theta)\, d\theta + \int_{\underline{\theta}}^{\overline{\theta}} [v(\theta) - x_i(\theta)] w(\theta)\, d\theta},$$

$$L_2(Index_i) = \frac{\sqrt{\int_{\underline{\theta}}^{\overline{\theta}} (x_i(\theta) - u(\theta))^2 w(\theta)\, d\theta}}{\sqrt{\int_{\underline{\theta}}^{\overline{\theta}} (x_i(\theta) - u(\theta))^2 w(\theta)\, d\theta} + \sqrt{\int_{\underline{\theta}}^{\overline{\theta}} (x_i(\theta) - v(\theta))^2 w(\theta)\, d\theta}},$$

$$i = 1, \dots, K.$$

3. Main Result

In this section we consider two standard loss functions, namely, absolute error loss (L_1-norm) and squared error loss (L_2-norm).

3.1 Analysis based on L_1-norm

We first prove a general result in the case of L_1-norm, showing that the MCDM approach in this case is equivalent to a standard Bayesian approach. Such a result does not hold under the L_2-norm. Suppose $T_1,\dots,T_K$ are some estimates of θ to be compared with respect to their mean squared errors (MSE) $\text{MSE}(T_i) = x_i(\theta),\ i=1,\dots,K,$ where $\underline{\theta} \le \theta \le \overline{\theta}$.

Theorem 3.1 : Under L_1-norm, T_i is better than T_j if

$$\int_{\underline{\theta}}^{\overline{\theta}} x_i(\theta)\, w(\theta)\, d\theta \;<\; \int_{\underline{\theta}}^{\overline{\theta}} x_j(\theta)\, w(\theta)\, d\theta . \qquad (3.1)$$

Proof : T_i is better than T_j if

$$L_1(Index_i) \;<\; L_1(Index_j)$$

$$\Leftrightarrow \qquad \frac{L_1(i, IDR)}{L_1(i, NIDR)} \;<\; \frac{L_1(j, IDR)}{L_1(j, NIDR)} .$$

$$\Leftrightarrow \qquad \frac{\int_{\underline{\theta}}^{\overline{\theta}} \left[x_i(\theta) - u(\theta) \right] w(\theta)\, d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \left[v(\theta) - x_i(\theta) \right] w(\theta)\, d\theta} \;<\; \frac{\int_{\underline{\theta}}^{\overline{\theta}} \left[x_j(\theta) - u(\theta) \right] w(\theta)\, d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \left[v(\theta) - x_j(\theta) \right] w(\theta)\, d\theta}$$

$\Leftrightarrow$

$$\left[\int_{\underline{\theta}}^{\overline{\theta}} x_i(\theta)\, w(\theta)\, d\theta - A \right] \left[B - \int_{\underline{\theta}}^{\overline{\theta}} x_j(\theta)\, w(\theta)\, d\theta \right] < \left[\int_{\underline{\theta}}^{\overline{\theta}} x_j(\theta)\, w(\theta)\, d\theta - A \right] \left[B - \int_{\underline{\theta}}^{\overline{\theta}} x_i(\theta)\, w(\theta)\, d\theta \right]$$

where

$$A = \int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, w(\theta)\, d\theta \quad \text{and} \quad B = \int_{\underline{\theta}}^{\overline{\theta}} v(\theta)\, w(\theta)\, d\theta .$$

Therefore

$$[B - A] \left[\int_{\underline{\theta}}^{\overline{\theta}} x_i(\theta)\, w(\theta)\, d\theta - \int_{\underline{\theta}}^{\overline{\theta}} x_j(\theta)\, w(\theta)\, d\theta \right] \;<\; 0 .$$

Since $[B - A] > 0$, thus T_i is better than T_j if

$$\int_{\underline{\theta}}^{\overline{\theta}} x_i(\theta)\, w(\theta)\, d\theta \quad < \quad \int_{\underline{\theta}}^{\overline{\theta}} x_j(\theta)\, w(\theta)\, d\theta .$$

This completes the proof.

Corollary 3.2 : Let θ be a binomial proportion, $0 < \theta < 1$. If the weight function is defined by $w_1(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$ with $\alpha = \beta = \sqrt{n}/2$, $T_2(x)=[x+\sqrt{n}/2]/[n+\sqrt{n}]$ is the best estimate of θ under the MCDM approach.

The robustness of $T_2(x)$ for some other choices of α and β can be seen from the following cases where we mention values of n for which $T_2(x)$ is better than $T_1(x)$.

Case 1 : $\alpha = \beta = 1 : n \in [1,19]$ **Case 2** : $\alpha = \beta = 1.5 : n \in [1,41]$.

Case 3 : $\alpha = \beta = 2 : n \in [1,71]$. **Case 4** : $\alpha = \beta = 3 : n \in [1,155]$.

Case 5 : $\alpha = 1,\ \beta = 2 : n \in [1,19]$. **Case 6** : $\alpha = 1.5,\ \beta = 3 : n \in [1,33]$.

Case 7 : $\alpha = 1,\ \beta = 4 : n \in [1,7]$. **Case 8** : $\alpha = 2,\ \beta = 3 : n \in [1,71]$.

Case 9 : $\alpha = 2,\ \beta = 4 : n \in [1,47]$.

Following Filar [4], we now consider two additional choices of $w(\theta)$. The first one, denoted by $w_2(\theta)$, is based on the notion of entropy between MSE(T_1) and MSE(T_2) for various values of θ, and the second one, denoted by $w_3(\theta)$, is based on the coefficient of variation of MSE(T_1) and MSE(T_2) for various values of θ.

In the context of binomial parameter estimation problem, recall that $X \sim B(n,\theta)$, $T_1(x)=x/n$, $T_2(x)=[x+\sqrt{n}/2]/[n+\sqrt{n}]$, MSE$(T_1) = \dfrac{\theta(1-\theta)}{n}$ and MSE(T_2) = $\dfrac{n}{4(n+\sqrt{n})^2}$. It then readily turns out that

$$w_2(\theta) = \frac{1-\phi(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} [1-\phi(\theta)]\, d\theta} \tag{3.2}$$

where

$$\phi(\theta) = -\frac{1}{\log 2}\left\{\left(\frac{\frac{\theta(1-\theta)}{n}}{\frac{\theta(1-\theta)}{n}+\frac{n}{4(n+\sqrt{n})^2}}\right)\cdot\log\left(\frac{\frac{\theta(1-\theta)}{n}}{\frac{\theta(1-\theta)}{n}+\frac{n}{4(n+\sqrt{n})^2}}\right)+\left(\frac{\frac{n}{4(n+\sqrt{n})^2}}{\frac{\theta(1-\theta)}{n}+\frac{n}{4(n+\sqrt{n})^2}}\right)\cdot\log\left(\frac{\frac{n}{4(n+\sqrt{n})^2}}{\frac{\theta(1-\theta)}{n}+\frac{n}{4(n+\sqrt{n})^2}}\right)\right\}$$

and

$$w_3(\theta) = \frac{\left|\frac{\theta(1-\theta)}{n}-\frac{n}{4(n+\sqrt{n})^2}\right|}{\frac{\theta(1-\theta)}{n}+\frac{n}{4(n+\sqrt{n})^2}}. \tag{3.3}$$

Verification of (3.1) for these two weight functions has been carried out for various values of n, using MATHEMATICA. The results are stated below.

Corollary 3.3 : Let θ be a binomial proportion, $0<\theta<1$. Under $w_2(\theta)$, $T_2(x)$ is better than $T_1(x)$ for all $n \geq 1$.

Corollary 3.4 : Let θ be a binomial proportion, $0<\theta<1$. Under $w_3(\theta)$, $T_2(x)$ is better than $T_1(x)$ for all $n \in [1,19]$.

3.2 Analysis based on L_2-norm

In the case of L_2-norm with a general weight function $w(\theta)$, proceeding as before, it is easily seen that T_i is better than T_j if

$$L_2(Index_i) < L_2(Index_j)$$

$$\Leftrightarrow \quad \frac{L_2(i,IDR)}{L_2(i,NIDR)} < \frac{L_2(j,IDR)}{L_2(j,NIDR)}.$$

$$\Leftrightarrow \quad \frac{\int_{\underline{\theta}}^{\overline{\theta}}[x_i(\theta)-u(\theta)]^2\, w(\theta)\,d\theta}{\int_{\underline{\theta}}^{\overline{\theta}}[x_i(\theta)-v(\theta)]^2\, w(\theta)\,d\theta} < \frac{\int_{\underline{\theta}}^{\overline{\theta}}[x_j(\theta)-u(\theta)]^2\, w(\theta)\,d\theta}{\int_{\underline{\theta}}^{\overline{\theta}}[x_j(\theta)-v(\theta)]^2\, w(\theta)\,d\theta}. \tag{3.4}$$

We now simplify (3.4) in the case of our problem of comparison of T_1 and T_2 for estimation of the binomial proportion θ. Obviously $\mathrm{MSE}(T_2) \leq \mathrm{MSE}(T_1)$ holds whenever $c_1(n) < \theta < c_2(n)$, where

$$c_1(n) = \frac{1 - \sqrt{1 - \left(\frac{n}{n+\sqrt{n}}\right)^2}}{2} \tag{3.5}$$

and

$$c_2(n) = \frac{1 + \sqrt{1 - \left(\frac{n}{n+\sqrt{n}}\right)^2}}{2} \tag{3.6}$$

Moreover, the Ideal row and Negative-ideal row are easily obtained as

$u(\theta): IDR =$

$$\left\{ \frac{\theta(1-\theta)}{n} : \theta < c_1(n)\,,\ \frac{n}{4(n+\sqrt{n})^2} : c_1(n) < \theta < c_2(n)\,,\ \frac{\theta(1-\theta)}{n} : \theta > c_2(n) \right\}$$

$v(\theta): NIDR =$

$$\left\{ \frac{n}{4(n+\sqrt{n})^2} : \theta < c_1(n)\,,\ \frac{\theta(1-\theta)}{n} : c_1(n) < \theta < c_2(n)\,,\ \frac{n}{4(n+\sqrt{n})^2} : \theta > c_2(n) \right\}.$$

For $i = 1$, applying equations (2.5) and (2.6), we get

$$L_2(1, IDR) = \sqrt{\int_{c_1(n)}^{c_2(n)} \left(\frac{\theta(1-\theta)}{n} - \frac{n}{4(n+\sqrt{n})^2} \right)^2 w(\theta)\, d\theta}$$

$$L_2(1, NIDR) = \sqrt{\int_0^{c_1(n)} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta)\, d\theta + \int_{c_2(n)}^{1} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta) d\theta}$$

For $i = 2$, applying equations (2.5) and (2.6), we obtain

$$L_2(2, IDR) = \sqrt{\int_0^{c_1(n)} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta)\, d\theta + \int_{c_2(n)}^{1} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta) d\theta}$$

$$L_2(2, NIDR) = \sqrt{\int_{c_1(n)}^{c_2(n)} \left(\frac{\theta(1-\theta)}{n} - \frac{n}{4(n+\sqrt{n})^2} \right)^2 w(\theta)\, d\theta} \,.$$

Therefore equation (3.4) is now simplified as

$$\int_{\theta<c_1(n)\cup\theta>c_2(n)} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta)\, d\theta < \int_{c_1(n)}^{c_2(n)} \left(\frac{\theta(1-\theta)}{n} - \frac{n}{4(n+\sqrt{n})^2} \right)^2 w(\theta) d\theta$$

$\Leftrightarrow$

$$\int_0^1 \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta)\, d\theta \; < \; 2 \int_{c_1(n)}^{c_2(n)} \left(\frac{n}{4(n+\sqrt{n})^2} - \frac{\theta(1-\theta)}{n} \right)^2 w(\theta)\, d\theta \,. \tag{3.7}$$

Theorem 3.5 : Let θ be a binomial proportion, $0 < \theta < 1$. If the weight function is defined by $w(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$ for some $\alpha, \beta > 0$, then $T_2(x)$ is better than $T_1(x)$ based on L_2-norm if n satisfies (3.8).

Proof : Since $w(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$, (3.7) reduces to

$$\int_0^1 \left(\frac{n^2}{16(n+\sqrt{n})^4} - \frac{\theta(1-\theta)}{2(n+\sqrt{n})^2} + \frac{\theta^2(1-\theta)^2}{n^2} \right) \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta$$

$$< \int_{c_1(n)}^{c_2(n)} \left(\frac{n^2}{8(n+\sqrt{n})^4} - \frac{\theta(1-\theta)}{(n+\sqrt{n})^2} + \frac{2\theta^2(1-\theta)^2}{n^2} \right) \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta$$

$\Leftrightarrow$

$$\frac{n^2}{16(n+\sqrt{n})^4} \cdot \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{1}{2(n+\sqrt{n})^2} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} + \frac{1}{n^2} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+4)}$$

$$< \int_{c_1(n)}^{c_2(n)} \left(\frac{n^2\theta^{\alpha-1}(1-\theta)^{\beta-1}}{8(n+\sqrt{n})^4} - \frac{\theta^{\alpha}(1-\theta)^{\beta}}{(n+\sqrt{n})^2} + \frac{2\theta^{\alpha+1}(1-\theta)^{\beta+1}}{n^2} \right) d\theta . \tag{3.8}$$

We now consider some special cases of α and β and indicate values of n for which (3.8) holds.

Case 1 : $\alpha = \beta = 1 : n \in [1,14]$. **Case 2** : $\alpha = \beta = 1.5 : n \in [1,26]$.

Case 3 : $\alpha = \beta = 2 : n \in [1,41]$. **Case 4 :** $\alpha = \beta = 3 : n \in [1,82]$.

Case 5 : $\alpha = \beta = \sqrt{n}/2 : n \in \mathrm{I}^+$. **Case 6** : $\alpha = 1,\ \beta = 2 : n \in [1,14]$.

Case 7 : $\alpha = 1.5,\ \beta = 3 : n \in [1,22]$. **Case 8** : $\alpha = 1,\ \beta = 4 : n \in [1,7]$.

Case 9 : $\alpha = 2,\ \beta = 3 : n \in [1,41]$. **Case 10 :** $\alpha = 2,\ \beta = 4 : n \in [1,30]$.

It is interesting to observe that in this case also, T_2 outperforms T_1 for all values of n under the minimax prior distribution.

As in the case of L_1-norm, here also we considered the other two weight functions $w_2(\theta)$ and $w_3(\theta)$, and verified (3.7) for various values of n, using MATHEMATICA. The results are stated below.

Corollary 3.6 : Let θ be a binomial proportion, $0 < \theta < 1$. If the weight function is defined by (3.2) then T_2 is better than T_1 based on L_2-norm for all $n \geq 1$.

Corollary 3.7 : Let θ be a binomial proportion, $0 < \theta < 1$. If the weight function is defined by (3.3) then T_2 is better than T_1 based on L_2-norm if $n \in [1,14]$.

4. Conclusion

Based on the above analysis under L_1- and L_2- norms, our recommendation is to use T_2 rather than T_1 for small values of n.

5. References

[1] Casella, G., and Berger, R.L. (1990). Statistical Inference. Wadsworth Inc. California, USA.

[2] Zeleny, M.(1982). Multiple Criteria Decision Making. McGraw-Hill, New York, USA.

[3] Lertprapai, S., Tiensuwan, M., and Sinha Bimal, K. (2003). *A statistical approach to combining environmental indices with an application to air pollution data from Bangkok, Thailand.*Technical Report, Department of Mathematics and Statistics,UMBC and Department of Mathematics, Mahidol University, Thailand.

[4] Filar, J.A., Ross, N.P., and Wu, M.L.(1999). Environmental Assessment Based on Multiple Indicators. Technical Report, Department of Applied Mathematics, University of South Australia, Australia.